

УДК 004.414

О.О. Івашко

Тернопільський національний технічний університет імені Івана Пулюя, Україна

## ПРИНЦИП РОБОТИ ПОШУКОВИХ МАШИН

О.О. Ivashko

### PRINCIPLE OF OPERATION SEARCH ENGINE

Механізми пошуку в інформаційно-пошуковій системі (ІПС) – це система, що забезпечує пошук і відбір необхідних даних у спеціальній базі з описами джерел інформації на основі інформаційно-пошукової мови і відповідних правил пошуку.

Пошукові системи включають три основні компоненти:

- веб-сторінку з пошуковим механізмом, яка виконує роль інтерфейсу для організації взаємодії з базою даних;
- базу даних, де міститься інформація, що зібрана спеціальними програмами пошукової системи. Власне наявністю баз даних пояснюється висока швидкість виведення результатів пошуку на сторінку пошукової системи;
- пошукові роботи (Robots), павуки (Spiders) або хробаки (Worms) - спеціальні програмні засоби, які автоматично періодично «відвідують» сайти, збирають відомості про вміст сторінок, тобто індексують їх і наповнюють бази даних пошукової системи.

Схема інформаційно-пошукової системи наведена на рисунку 1 [1].

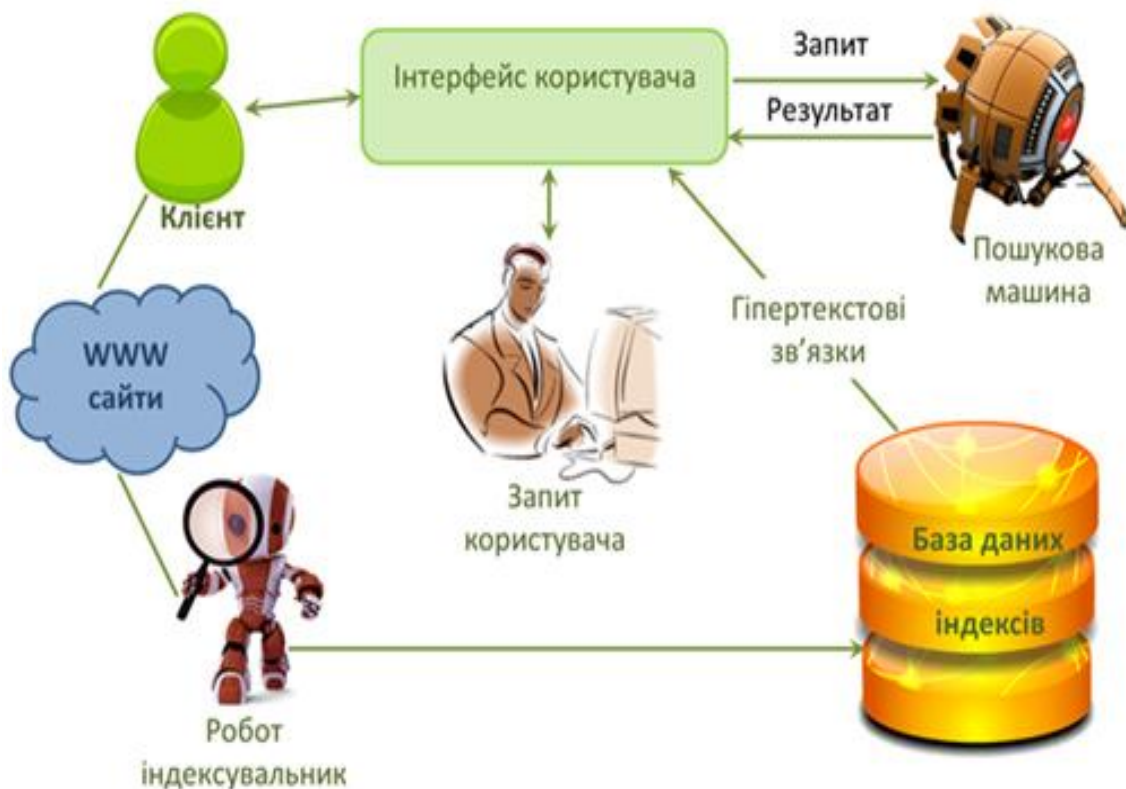


Рис. 1. Типова схема інформаційно-пошукової системи

Більшість пошукових систем світу - індексні пошукові системи, які ще називають пошуковими покажчиками, пошуковими серверами, словниковими пошуковими системами, автоматичними індексами, пошуковими машинами, Search

Engines тощо. Їхнє призначення – якнайкраще охопити інформаційних веб-простір і подати його користувачам у зручному вигляді.

В основу принципу роботи з індексними пошуковими системами (ІПС) покладено ключові слова [2]. Починаючи пошук інформації з певної теми, користувач має ввести ключові слова у рядок пошуку будь-якого браузера, які описують його тему і вибрати кнопку Знайти, яка розташована поряд з рядком пошуку.

Після вибору кнопки Знайти, ключова фраза посиляється на сервер і пошукова система починає пошук у своїй базі даних адреси веб-ресурсів, які містять вказані ключові слова. Як здійснюється пошук, користувач не може бачити, він лише побачить результат запиту – нове вікно, в якому відображається перелік веб-ресурсів, які містять ключові слова.

Пошукова система враховує такі параметри сайту при обчисленні його релевантності:

- частота ключових слів: складні алгоритми сучасних пошукачів дозволяють проводити семантичний аналіз тексту, щоб відсіяти пошуковий спам, коли ключове слово зустрічається дуже часто або дуже рідко на сторінці. Оптимальна щільність ключових слів: 5-7 % від текстового вмісту веб-сторінки, при цьому ключові слова повинні зустрічатися в тексті як мінімум 3-4 рази;

- індекс цитування сайту, або кількість веб-ресурсів, що посиляються на даний сайт; багатьма пошукачами не враховуються взаємні посилання (один на одного), також важливо, щоб посилання були з веб-ресурсів схожої тематики.

Відповідно, всі чинники, що впливають на положення сайту у видачі пошукової системи, можна поділити на зовнішні і внутрішні. Робота по оптимізації включає роботу:

- з внутрішніми чинниками які знаходяться під контролем власника веб-сайту – приведення тексту і розмітки сторінок у відповідність з вибраними запитами, поліпшення якості і кількості тексту на сайті, стилістичне оформлення тексту, поліпшення структури і навігації, використання внутрішніх посилань;

- зовнішніми чинниками – обмін посиланнями, реєстрація в каталогах та інші заходи для підвищення і стимулювання кількості і частоти посилання на ресурс.

### **Література**

1. Інформаційно-пошукові системи // Соціальна інформатика та технології. – Режим доступу: [http://abramchuk-inf.blogspot.com/p/blog-page\\_6023.html](http://abramchuk-inf.blogspot.com/p/blog-page_6023.html). – Дата доступу: жовтень 2015 року. – Заголовок з екрану.

2. Волощук Николай. Інформаційно-пошукова система / Николай Волощук // Экономика ИТ-бизнеса. – Режим доступу: [http://forstudents.at.ua/publ/ehkonomika\\_it\\_biznesa/ehkonomika\\_it\\_biznesa/informacijno\\_po\\_shukova\\_sistema/73-1-0-897](http://forstudents.at.ua/publ/ehkonomika_it_biznesa/ehkonomika_it_biznesa/informacijno_po_shukova_sistema/73-1-0-897). – Дата доступу: жовтень 2015 року. – Заголовок з екрану.